LA-UR-92-2972

(Conf-9208161--)

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

LA-UR--92-2972

DE93 000687

TITLE: Statistical error in a chord estimator of correlation dimension: the "rule of five"

AUTHOR(S): James Theiler, Los Alamos National Laboratory
Turab Lookman, University of Western Ontario

SUBMITTED TO: International Journal of Bifurcations & Chaos
Proceedings of Workshop on Dynamical Measures of Complexity and Chaos II (Bryn Mawr, 13-15 Aug '92)

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Statistical error in a chord estimator of correlation dimension: the "rule of five"

James Theiler
Center for Nonlinear Studies and Theoretical Division,
Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Turab Lookman
Department of Applied Mathematics,
University of Western Ontario, London, Ontario N6A 5B7, CANADA

(Revised September 1, 1992)

ABSTRACT

The statistical precision of a chord method for estimating dimension from a correlation integral is derived. The optimal chord length is determined, and a comparison is made to other estimators. The simple chord estimator is only 25% less precise than the optimal estimator which uses the full resolution and full range of the correlation integral. The analytic calculations are based on the hypothesis that all pairwise distances between the points in the embedding space are statistically independent. The adequacy of this approximation is assessed numerically, and a surprising result is observed in which dimension estimators can be anomalously precise for sets with reasonably uniform (nonfractal) distributions.

Running title: a chord estimator of correlation dimension
Keywords: correlation dimension, statistical error
PACS: 02.60.+y, 05.45.+b

1. Introduction

Estimating the fractal dimension of a strange attractor from a finite sample of points is a problem that has attracted considerable interest over the last decade, particularly in the context of time series analysis.[1] Insofar as dimension can be accurately computed, it provides a very useful measure of the underlying complexity of the dynamics by counting the number of "active" degrees of freedom.

A standard approach for estimating the dimension involves first reconstructing the phase space by embedding the data in a higher dimensional space; this generally involves the use of time delay coordinates [Packard *et al.*, 1980] or their linear combinations [Broomhead

[1] See Tong [1990], Grassberger *et al.* [1991], and Casdagli *et al.* [1992] for recent reviews of time series in general, and Theiler [1990a] for a review of dimension estimation in particular.

and King, 1987]. The issues involved in making a good embedding are beyond the scope of this brief article; see Sauer *et al.* [1991], Casdagli *et al.* [1992b], and Gibson *et al.* [1992a] for recent expositions. For our purposes, it will be assumed that we have a sample of N independent data points in an m dimensional phase space: $\{x_1, \ldots, x_n\} \in \mathbf{R}^m$.

One can define a *pointwise* mass function $B(x; N, r)$ which counts the fraction of points (not including the reference point x) that are within a distance r of the point x. The scaling of $B(x; N, r)$ with r for large N and small r defines a pointwise dimension: $B(x; N, r) \sim R^{d(x)}$.

Averaging the pointwise mass function over all the points x leads to the correlation integral of Grassberger and Procaccia [1983] and Takens [1983]. Here $C(N, r) = \langle B(x; N, r) \rangle_x$ is the fraction of distances smaller than r between all pairs of points in an n point sample.[2] If, in the limit of large N and small r, the correlation integral scales as r^ν, then the exponent ν defines the correlation dimension. Formally,

$$\nu = \lim_{r \to 0} \lim_{N \to \infty} \frac{\log C(N, r)}{\log r} \tag{1}$$

In practice, ν is usually estimated as the slope in a log-log plot of $C(N, r)$ versus r, restricted to some range of r over which "good scaling" is observed. While there is still a fair bit of art to choosing this scaling range, the actual fit of the slope is fairly straightforward. It is the fitting of this slope that will be addressed here.

One very common approach is to estimate the slope by taking a least squares fit through the points along some segments on the $C(N, r)$ curve. This is a reasonable approach, as long as it is done properly [Denker and Keller, 1986; Cutler, 1991], but the method is often abused. Some authors are led to believe that it is optimal because least squares was employed. Even more dangerous is a tendency to associate the standard error of the unweighted least-squares fit with the "error bar" on the dimension itself; the two are utterly different.

1.1 Takens Estimator

Takens [1985] proposed a "best" estimator for ν directly from the set of pairwise distances. (Smith [1992a] pointed out that the same estimator was derived much earlier in another context, before being "rediscovered" by Takens, though Takens was the first to apply it to the problem of estimating the dimension of a chaotic attractor.) The Takens estimator requires the choice of a single free parameter, R_0, the upper cutoff distance. All pairwise distances larger than R_0 are discarded, and all distances r which are less than R_0 are averaged according to

$$\hat{\nu} = \frac{-1}{\langle \log(r/R_0) \rangle} \tag{2}$$

The Takens estimator is optimal in the situation that all distances are independent (which is not strictly true, but is often a good hypothesis for small distances — more about this later),

[2] Efficient implementations are described by Theiler [1987] and Grassberger [1990].

and the correlation integral is strictly proportional to r^ν. [3] See Cawley and Licht [1986] for some numerical experiments with this estimator.

An equivalent form of the Takens estimator, in terms of the correlation integral, is given by

$$\hat{\nu} = \frac{C(R_o)}{\int_0^{R_o}[C(r)/r]\,dr} \tag{3}$$

1.2 Chord Estimator

Arguably the easiest way to estimate a slope of the correlation integral to choose two points on the curve, at say R_o and R_1, and measure the slope of the chord that is drawn through those two points. That is,

$$\hat{\nu} = \frac{\log C(N, R_o) - \log C(N, R_1)}{\log R_o - \log R_1} \tag{4}$$

This may seem terribly inefficient, because so much information is thrown away, but for a well chosen chord, the statistical error is only be 25% worse than for the optimal Takens estimator, which uses all the information in the $C(N, r)$ function.

In choosing the optimal chord, two effects are traded off against each other. First, the "length" of the chord should be as large as possible, so that the slope is less sensitive to small fluctuations in the endpoints. Second, the chord should not be so long as to reach too far into the small distances, where the fluctuations of the endpoint is particularly large. In what follows, the optimal choice of chord will be derived; the rule for optimal choice can be expressed independently of ν and of the upper cutoff R_o.

Because Eq. (4) provides a simple and explicit expression for estimated dimension, a number of authors have used it for more general calculations which attempt to estimate the number of data points required to achieve good dimension calculations [Nerenberg and Essex, 1990; Essex and Nerenberg, 1992; Majski and Lookman, 1992].

2. Derivation of Statistical Error and Optimal Chord

Suppose the correlation integral is evaluated at two distances, R_o and $R_1 < R_o$. Define $\mathcal{R} = R_o/R_1 > 1$. Let $\mathcal{N}_o$ and $\mathcal{N}_1$ be the number of distances less than R_o and R_1 respectively. Finally define $n_o = \mathcal{N}_o - \mathcal{N}_1$. The chord estimate of dimension is given by

$$\hat{\nu} = \frac{\log \mathcal{N}_o - \log \mathcal{N}_1}{\log R_o - \log R_1} = \frac{\log(\mathcal{N}_1 + n_o) - \log \mathcal{N}_1}{\log \mathcal{R}} \tag{5}$$

Now, $\mathcal{N}_1$ and $\mathcal{N}_o$ are not statistically independent, but $\mathcal{N}_1$ and $n_o = \mathcal{N}_o - \mathcal{N}_1$ are. Further, since the random variables $\mathcal{N}_1$ and n_o have a poisson distribution, we can write the variances

[3] This strict proportionality can fail because of noise in the data, boundary effects, or "lacunarity." The last of these has been described by a number of authors; a recent and comprehensive treatment of the issue as it applies to dimension estimation is provided by Smith [1992c].

$\mathrm{Var}(\mathcal{N}_1) = \mathcal{N}_1$ and $\mathrm{Var}(n_o) = n_o$. Then, we can compute

$$\mathrm{Var}(\nu) = \left(\frac{\partial \nu}{\partial \mathcal{N}_1}\right)^2 \mathrm{Var}(\mathcal{N}_1) + \left(\frac{\partial \nu}{\partial n_o}\right)^2 \mathrm{Var}(n_o). \tag{6}$$

In particular,

$$\frac{\partial \nu}{\partial \mathcal{N}_1} = \frac{1}{\log \mathcal{R}} \left[\frac{1}{\mathcal{N}_1 + n_o} - \frac{1}{\mathcal{N}_1}\right] \tag{7}$$

$$\frac{\partial \nu}{\partial n_o} = \frac{1}{\log \mathcal{R}} \left[\frac{1}{\mathcal{N}_1 + n_o}\right]. \tag{8}$$

Thus, we can write

$$\mathrm{Var}(\nu) = \frac{1}{(\log \mathcal{R})^2} \left(\left[\frac{1}{\mathcal{N}_1 + n_o} - \frac{1}{\mathcal{N}_1}\right]^2 \mathcal{N}_1 + \left[\frac{1}{\mathcal{N}_1 + n_o}\right]^2 n_o \right). \tag{9}$$

Let us introduce the parameter Θ defined by

$$\Theta = \frac{C(N, R_o)}{C(N, R_1)} = \mathcal{N}_o / \mathcal{N}_1 = (R_o / R_1)^\nu. \tag{10}$$

Then, use $\mathcal{N}_1 = \mathcal{N}_o / \Theta$ and $n_o = \mathcal{N}_o(\Theta - 1)/\Theta$, as well as $\nu \log \mathcal{R} = \log \Theta$ to rewrite Eq. (9).

$$\mathrm{Var}(\nu)/\nu^2 = \frac{1}{\mathcal{N}_o} \left[\frac{\Theta - 1}{(\log \Theta)^2}\right]. \tag{11}$$

This expression is independent of ν and of R_o, and is minimized with Θ approximately equal to 4.921554, giving a relative statistical error for ν as about $1.25/\sqrt{\mathcal{N}_o}$. By contrast, the Takens best estimator (which uses the correlation integral at *all* distances less than the upper cutoff R_o instead of just the single distance R_1) has a relative error of $1.0/\sqrt{\mathcal{N}_o}$.

Having chosen an upper scale R_o, the optimal choice of lower scale R_1 is given by that R_1 for which

$$C(R_o)/C(R_1) = \Theta_{\mathrm{optimal}} \approx 5. \tag{12}$$

Using this "rule of five," it is possible to very rapidly estimate a fractal dimension which is only slightly less efficient than optimal.

Note that there is no penalty for large dimension. This analysis also suggests that the optimal choice of R_o is as large as possible. This is true from the point of view of statistical precision, but if systematic errors are included, smaller R_o are generally desired. If lacunarity is taken into account, then there is an advantage to larger $\mathcal{R}$.

3. Comparison to Other Estimators

The Takens method uses all distances less than R_o, whereas the chord method uses $C(N,r)$ at only two values of r. Qualitatively speaking, there are two different sources of the increased imprecison in the chord method. One results from ignoring distances below the lower cutoff R_1, and the other results from ignoring distances between R_1 and R_o. It is possible to assess the relative contributions of these two effects by considering two other estimators. The first was suggested by Ellner [1988] (see also Olofsen *et al.* [1992] for an extension to multiple embedding dimensions) who derived the maximum likelihood estimator that uses all distances between R_1 and R_o. Here,

$$\hat{\nu} = \frac{-1}{\frac{\mathcal{N}_1}{\mathcal{N}_o - \mathcal{N}_1}\log(R_1/R_o) + \langle \log(r/R_o)\rangle} \tag{13}$$

where the average is over all $\mathcal{N}_o - \mathcal{N}_1$ distances between R_1 and R_o. Note that as $R_1 \to 0$, $\mathcal{N}_1/\mathcal{N}_o \to 0$ and the original Takens estimator is retrieved. In terms of the correlation integral, this estimator can be written

$$\hat{\nu} = \frac{C(R_o) - C(R_1)}{\int_{R_1}^{R_o}[C(r)/r]\,dr}. \tag{14}$$

Ellner [1988] has shown that the variance of this estimator is given by

$$\mathrm{Var}(\nu)/\nu^2 = \frac{1/[1-(R_1/R_o)^\nu]}{\mathcal{N}_o} \tag{15}$$

$$= \frac{1}{\mathcal{N}_o}\left[\frac{\Theta}{\Theta - 1}\right] \tag{16}$$

In particular, at $\Theta = 5$, the relative statistical error is $1.12/\sqrt{\mathcal{N}_o}$.

A second alternative estimator was suggested by Smith [1992b]; it is the maximum likelihood estimator based on $K+2$ distances equally spaced on a logarithmic scale: $R_K = R_o/\mathcal{R}^K$ (note that the special case $K = 0$ is the chord estimator):

$$\hat{\nu} = \frac{\log[C(R_o) + C(R_1) + \cdots + C(R_K)] - \log[C(R_1) + \cdots + C(R_{K+1})]}{\log \mathcal{R}} \tag{17}$$

It is interesting to consider the $K \to \infty$ limit for this estimator, because in opposition to Ellner's estimator the estimator does *not* throw away information in $C(N,r)$ for $r < R_1$, but *does* throw away information in $C(N,r)$ for distances between R_o and R_1. Ellner's estimator arguably makes more sense, because there may be good reason to throw away the $r < R_1$ information (noise may have corrupted the small distance scales), whereas the only case for ignoring $C(N,r)$ at the intermediate distance scales is ease of implementation (though this is not a trivial consideration). Smith's estimator is illustrative because it does well just when Ellner's does poorly (at small Θ), and vice versa (for large Θ).

For large K, it can be shown that

$$\mathrm{Var}(\nu)/\nu^2 = \left[\frac{(\Theta-1)^2}{\Theta(\log\Theta)^2}\right]\frac{1}{\mathcal{N}_o} \tag{18}$$

In the limit $\Theta \to 1$, this approaches $\mathrm{Var}(\nu)/\nu^2 = 1/\mathcal{N}_o$, which is the same as the Takens estimator. In fact, in the $\Theta \to 1$ limit, the $K \to \infty$ version of Eq. (17) is exactly equivalent to the Takens estimator. (If $\Theta \to 1$ and $K \to \infty$ in such a way that $K \log\Theta$ is fixed, then Ellner's estimator is approached.)

We should remark that the estimator in Eq. (17) is not Smith's *only* estimator. He has also introduced modifications to account for noise [Smith, 1992b] and lacunarity [Smith, 1992c].

4. The Independent Distance Hypothesis

All of the estimators described so far were derived to be optimal under the assumption that all distances less than R_o are independent. Although this is basically valid for pointwise dimension estimators, it is not true for the correlation dimension. One possibility is to randomly choose $N/2$ pairs using all N data points; as long as the data points are independent, so will the pairs be. This may be philosophically correct (PC), but it is very inefficient because it ignores most of the short distances.[4] If there are N independent data points, each with m independent coordinates, one might expect that the $O(mN)$ shortest distances are "nearly" independent; though they undoubtedly will not be strictly independent since they will be constrained, for instance, by various triangle inequalities.[5] Smith [1992b] provies both an informal [Smith, 1992b] and a more formal [Smith, 1992c] argument for the near-independence of the shortest distances.

In Fig. 2, we plot statistical error in a chord estimator of dimension for data uniformly distributed on the unit segment. For small R_o, we find that this error agrees with the prediction of Eq. (11), which assumes that distances are independent. For larger R_o, however, we find that the statistical error is larger than predicted by Eq. (11), at least for smaller Θ. (For larger Θ, the error is dominated by statistical error of $C(N, R_1)$, where R_1 is small, so distances less than R_1 are nearly independent; thus Eq. (11) predicts error at large Θ well even when R_o is large.) When R_o is large enough that the independent distances hypothesis leads to a poor approximation, the optimal Θ becomes larger than five.

[4] Computing a single pointwise dimension has the same problem: too many small distances are ignored. Some have suggested computing a correlation dimension based on distances from only a few references points, but the case against this is presented in [Theiler 1990b].

[5] Another possibility involves constructing a spanning tree, and computing distances only between points linked on the tree; this is a kind of compromise which sidesteps the triangle inequality constraints, and still provides mostly small distances. But it would still be incorrect to consider those distances as truly independent

4.1 Anomalous Precision in Dimension Estimation

It was shown in [Theiler, 1990b] that except for very special cases, the precision with which the correlation integral $C(N,r)$ is estimated scales as $1/\sqrt{N}$ for sufficiently large N. It is natural to assume that dimension estimates based on the correlation integral should similarly scale. But some recent numerical observations suggests that the estimated dimension for some attractors may have a precision which scales as $1/N$, even though the correlation integral itself is (true to theory) scaling as $1/\sqrt{N}$.

The effect is illustrated in Fig. 3. Two different data sets were generated, both randomly, but according to different distributions. In Fig. 3(ab), the statistical error for the correlation integral and the chord dimension are both shown as a function of N for data produced by a unit-variance gaussian random number generator. The correlation integral is evaluated at $R_o = 0.1$, and the chord dimension is estimated with $\Theta = 5$. Although the error in the correlation integral scales as $1/\sqrt{N}$, the error in the chord dimension appears to scale as $1/N$.

That this is not a general property is shown in Fig. 3(cd). Here data was generated using an iterated function system [Barnsley and Demko, 1985]: $x_{n+1} = (1/4)x_n$ with probability $1/2$; and $x_{n+1} = (3/4)x_n + (1/4)$ with probability $1/2$. Finally, we take only every tenth point (x_1, x_{11}, *etc.*), so that the points are effectively independent. The data fills the unit segment [0,1], but with a highly nonuniform Cantor-set-like measure.[6] The correlation dimension satisfies $4^\nu + (4/3)^\nu = 4$, for which $\nu = 0.733644$ is an approximate solution. Again the correlation integral and chord dimension is estimated at $R_o = 0.1$ and $R_1 = R_o/9$, so that $(R_o/R_1)^\nu \approx 5$. What we see in this case is that estimators for both the correlation integral and the correlation dimension have errors that scale like $1/\sqrt{N}$.

This anomalously precise estimate of dimension for sets with smooth integer-dimensional measures may account for the observation in [Theiler *et al.*, 1992 (Fig. 2a)] that the significance of a dimension-based test for nonlinearity in a low-dimensional time series of length N scales linearly with N. In that case, the given time series is compared surrogate data sets which are linear correlated noise (surrogate data) with the same autocorrelation as the original data. The significance is defined by the difference between the estimated dimension for the real and the surrogate data, divided by the standard deviation of the the dimension estimates for the many surrogate data sets. Since as we have seen, this standard deviation scales as $1/N$, this explains why the significance increases so rapidly for N.

One (still uninvestigated) implication of this anomalous scaling relates to the popular and increasingly contentious studies of data requirements. Attempts to estimate how much data one needs to compute dimension to a pre-specified accuracy are often based on quantifying the tradeoff between statistical error and boundary effects. This obviously depends on the nature of the underlying attractor, but if smooth nonfractal sets are used as "typical"

[6] The unit segment [0,1] has half of its measure concentrated in the interval [0,1/4], and the other half in [1/4,1]; each of these two intervals has half of its measure concentrated in the first quarter of its length, and the other half in the last three quarters; the measure in each of these four intervals is similarly partitioned, and so on.

examples, one may find that the statistical error is much smaller than actually occurs in the truly typical cases, and one may therefore underestimate the amount of data that is really required.

5. Reprise

Assuming the independent distances hypothesis, that all distances are known to full accuracy, and that the correlation integral scales as a strict power law, then the most precise estimator of the exponent of that power law ig given by Takens [1985]. By throwing away information in the $C(N,r)$ curve at distances less than a lower cutoff R_1 (as one might do if these distances are contaminated by noise), the best estimator is given by Ellner [1988]. For very small R_1 (large Θ), not much information is lost by ignoring these small distances, but for large $R_1 \approx R_o$, the effect of a small lever-arm is evident, and the estimator has a large statistical error.

If on the other hand one only knows the correlation integral $C(N,r)$ at discrete values of r that are equally spaced on a logarithmic scale, then the estimator of Smith [1992b] is best. in the $K \to \infty$ limit, it keeps information that Ellner's estimator discards, namely the behavior of $C(N,r)$ in the small r limit. It also discards information that Ellner's estimator uses in the behavior of $C(N,r)$ for distances between R_1 and R_o. This estimator works best when $R_1 \approx R_o$ (small Θ), because less information is lost.

The chord estimator is less precise than either Smith's or Ellner's estimator, because it considers only two points, and ignores everything else. However, when those two points are chosen so that the ratio of their correlation integrals is about five, then a very rapid estimate of dimension is obtained which is only 25% less precise than the optimal Takens estimate that uses the full resolution and full range of the correlation integral.

On the other hand, one does not expect the independent distances hypothesis to be valid; it is impossible to get $O(N^2)$ independent degrees of freedom from N points. Numerical experiments indicate that this hypothesis is reasonable if only the smallest $O(N)$ distances are used. But for certain nonfractal attractors, we have observed that the precision of a dimension estimator scales as $1/N$ with the number of data points, *as if* the independent distances hypothesis were applicable. The authors are presently unable to explain this anomalous effect.

Acknowledgements

We are grateful to the Advanced Computing Laboratory at the Los Alamos National Laboratory for computer resources used to perform the numerical experiments shown here. Work by JT was partially supported by a grant from the National Institute of Mental Health 1-R01-MH47184-01 and was performed at Los Alamos under the auspices of the Department of Energy.

References

Barnsley, M. F. & Demko, S. [1985] "Iterated function systems and the global construction of fractals," *Proc. R. Soc. Lond. A* **399**, 243.

Broomhead, D. S. & King, G. P. [1987] "Extracting qualitative dynamics from experimental data," *Physica D* **20**, 217.

Casdagli, M., Des Jardins, D., Eubank, S., Farmer, J. D., Gibson, J., Hunter, N. & Theiler, J. [1992] "Nonlinear modeling of chaotic time series: theory and applications," in *Applied Chaos* (Addison-Wesley) pp. 335–382.

Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. [1991] "State space reconstruction in the presence of noise," *Physica D* **51**, 52–98.

Cawley, R. & Licht, A. L. [1986] "Maximum likelihood method for evaluating correlation dimension," in *The Physics of Phase Space*, ed. Kim, Y. S. & Zachary, W. W., vol. 278 of *Lecture Notes in Physics* (Springer-Verlag, Berlin) pp. 90–103.

Cutler, C. D. [1991] "Some results on the behavior and estimation of the fractal dimensions of distributions on attractors," *J. Stat. Phys.* **62**, 651–708.

Denker, M. & Keller, G. [1986] "Rigorous statistical procedures for data from dynamical systems," *J. Stat. Phys.* **44**, 67–93.

Ellner, S. [1988] "Estimating attractor dimensions from limited data: a new method, with error estimates," *Phys. Lett. A* **133**, 128–133.

Essex, C. & Nerenberg, M. A. H. [1991] "Comments on 'Deterministic chaos: the science and the fiction' by D. Ruelle," *Proc. R. Soc. Lond. A* **435**, 287–292.

Gibson, J. F., Farmer, J. D., Casdagli, M. & Eubank, S. [1992] "An analytic approach to practical state space reconstruction," *Physica D* **57**, 1–30.

Grassberger, P. [1990] "An optimized box-assisted algorithm for fractal dimensions," *Phys. Lett. A* **148**, 63–68.

Grassberger, P. & Procaccia, I. [1983] "Measuring the strangeness of strange attractors," *Physica D* **9**, 189–208.

Grassberger, P., Schreiber, T. & Schaffrath, C. [1991] "Non-linear time sequence analysis," *Int. J. Bif. Chaos* **1**, 521–547.

Majski, W. & Lookman, T. [1992] "Data requirements for correlation dimension estimates," Preprint.

Nerenberg, M. A. H. & Essex, C. [1990] "Correlation dimension and systematic geometric effects," *Phys. Rev. A* **42**, 7065–7074.

Olofsen, E., Degoede, J. & Heijungs, R. [1992] "A maximum likelihood approach to correlation dimension and entropy estimation," *Bull. Math. Bio.* **54**, 45–58.

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Sauer, T., Yorke, J. A. & Casdagli, M. [1991] "Embedology," *J. Stat. Phys.* **65**, 579–616.

Smith, R. L. [1992a] "Comment: Relation between statistics and chaos," *Stat. Sci.* **7**, 109–113.

Smith, R. L. [1992b] "Estimating dimension in noisy chaotic time series," *J. R. Stat. Soc. B* **54**, 329–352.

Smith, R. L. [1992c] "Optimal estimation of fractal dimension," in *Nonlinear Modeling and Forecasting*, ed. Casdagli, M. & Eubank, S., vol. XII of *SFI Studies in the Sciences of Complexity* (Addison-Wesley) pp. 115–136.

Takens, F. [1983] "Invariants related to dimension and entropy," in *Atas do 13°*, Colóqkio Brasiliero de Matemática (Rio de Janeiro).

Takens, F. [1985] "On the numerical determination of the dimension of an attractor," in *Dynamical Systems and Bifurcations, Groningen, 1984*, ed. Braaksma, B. L. J., Broer, H. W. & Takens, F., vol. 1125 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin) pp. 99–106.

Theiler, J. [1987] "Efficient algorithm for estimating the correlation dimension from a set of discrete points," *Phys. Rev. A* **36**, 4456–4462.

Theiler, J. [1990a] "Estimating fractal dimension," *J. Opt. Soc. Am. A* **7**, 1055–1073.

Theiler, J. [1990b] "Statistical precision of dimension estimators," *Phys. Rev. A* **41**, 3038–3051.

Theiler, J., Galdrikian, B., Longtin, A., Eubank, S. & Farmer, J. D. [1992] "Using surrogate data to detect nonlinearity in time series," in *Nonlinear Modeling and Forecasting*, ed. Casdagli, M. & Eubank, S., vol. XII of *SFI Studies in the Sciences of Complexity* (Addison-Wesley) pp. 163–188.

Tong, H. [1990] *Non-linear Time Series: A Dynamical System Approach* (Clarendon Press, Oxford).

Figure captions

Figure 1. Coefficient of relative statistical error, defined by $\sqrt{\mathcal{N}_o \mathrm{Var}(\nu)}/\nu$, for four estimators as a function of $\Theta = C(N, R_o)/C(N, R_1) = (R_o/R_1)^\nu$. The relative statistical error itself is a factor of $1/\sqrt{\mathcal{N}_o}$ smaller than these curves. The solid line is the chord estimator, which estimates dimension from the $C(N, r)$ curve at only two values of r. The optimum chord estimator is given at $\Theta \approx 5$. At this value, it is 25% worse than the Takens estimator (dashed line), and about 11% worse than the estimators of Ellner (dotted line) and Smith (dashed dotted line). Note that the minimum is quite shallow; for $1.35 \leq \Theta \leq 75$ the statistical error is still no more than twice the optimal Takens error.

Figure 2. Theoretical (dashed line) and numerically estimated (solid line) statistical error for the chord estimator using data generated randomly and uniformly on the unit segment. In all cases $N = 300$ points were used; the three curves correspond (from top to bottom) to $R_o = 0.02, 0.10, 0.50$. The agreement with theory breaks down for large R_o, presumably because the approximation of independent distances fails.

Figure 3. Anomalous scaling of statistical error as a function of the number N of points in the data set. Panels (a,c) are for the correlation integral itself, evaluated at $R_o = 0.1$ for (a) unit variance gaussian noise, and (b) random data with a highly nonuniform Cantor-like distribution. In agreement with theory, both scale as $1/\sqrt{N}$. Panels (b,d) are for the estimate of chord dimension using $R_o = 0.1$ and $\Theta = 5$. Here, we observe the anomalous $1/N$ scaling for the gaussian distribution, while we see the expected $1/\sqrt{N}$ scaling for the Cantor distribution.

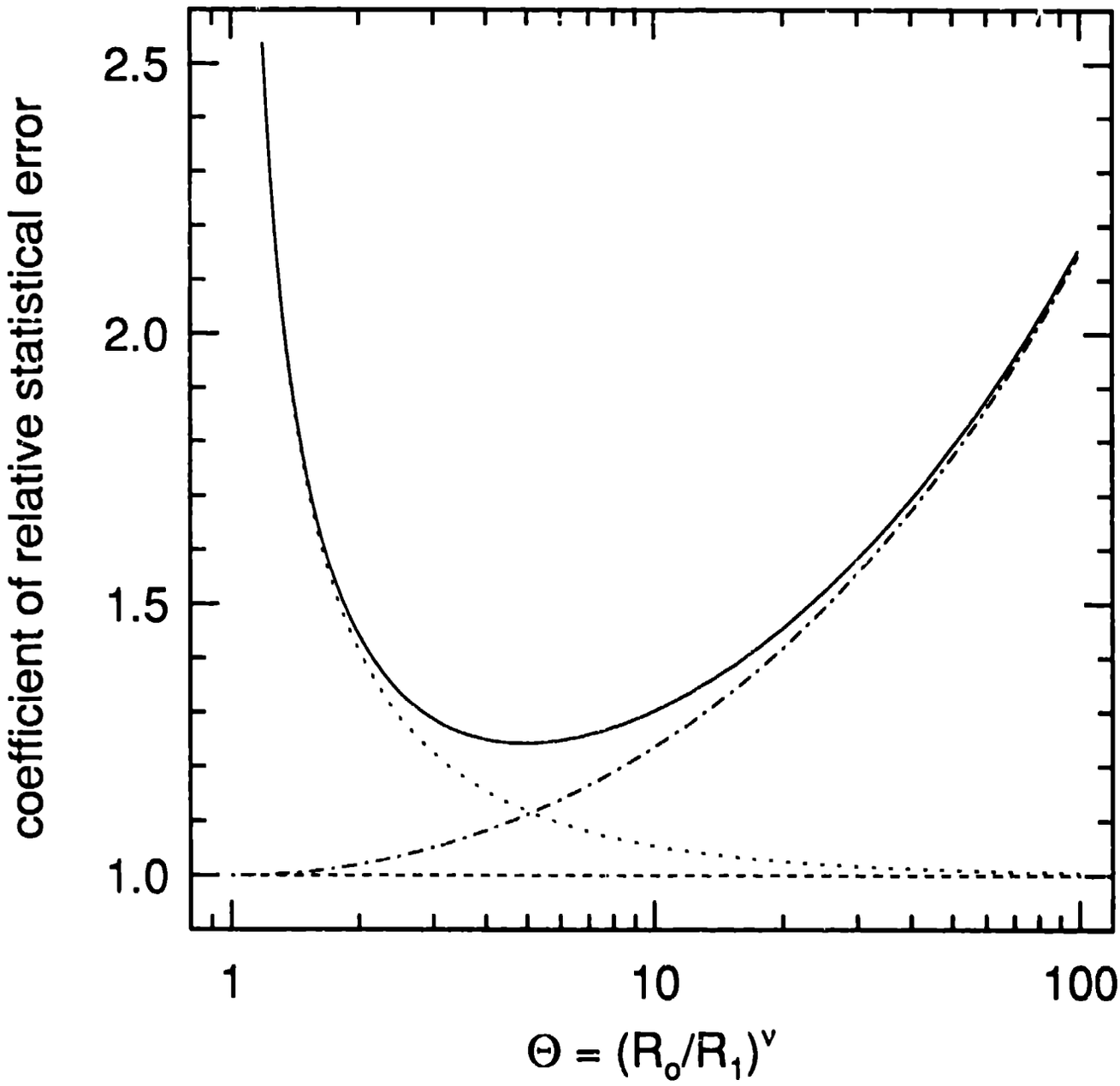

Figure 1: Theiler and Lookman

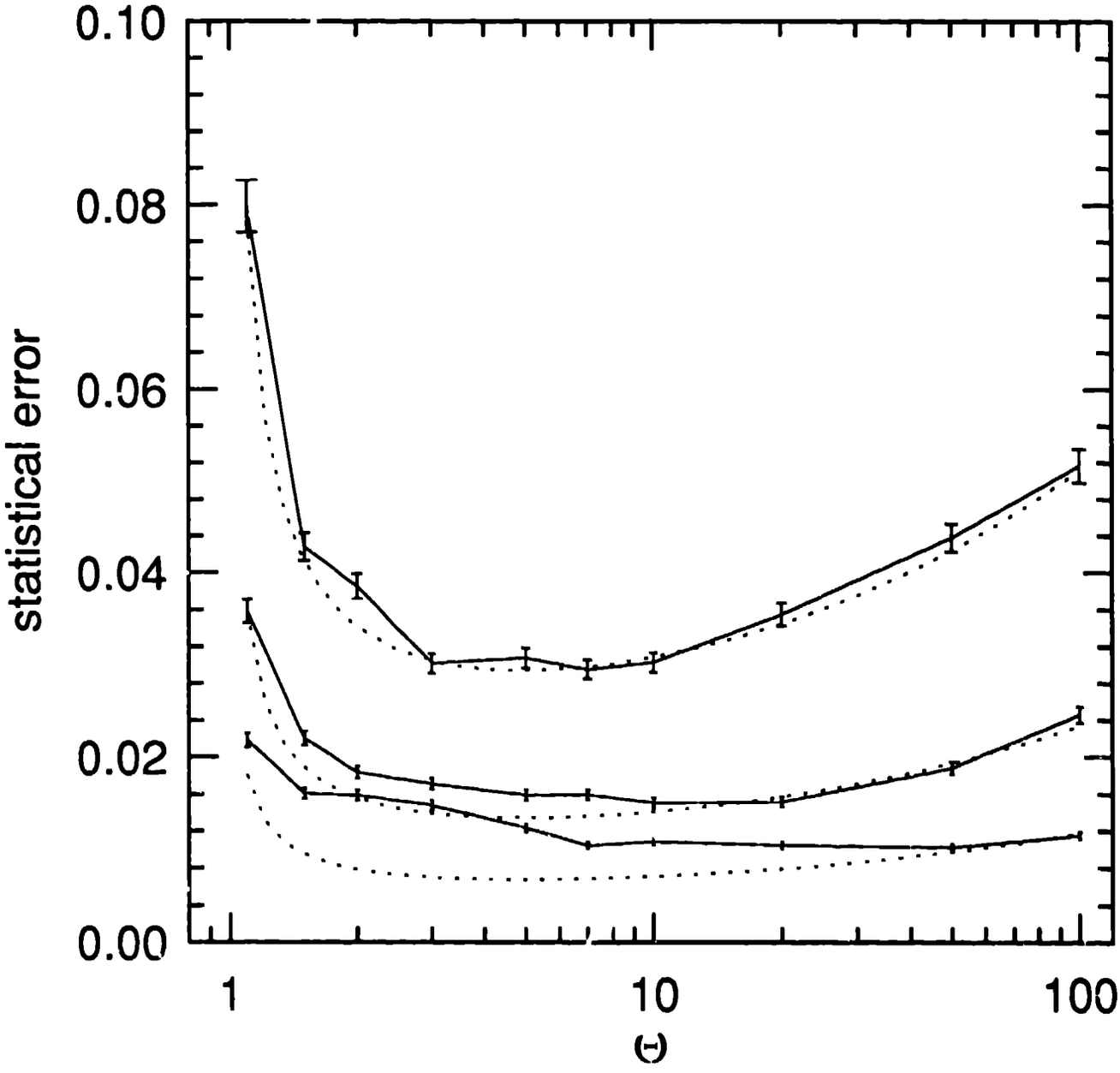

Figure 2: Theiler and Lookman

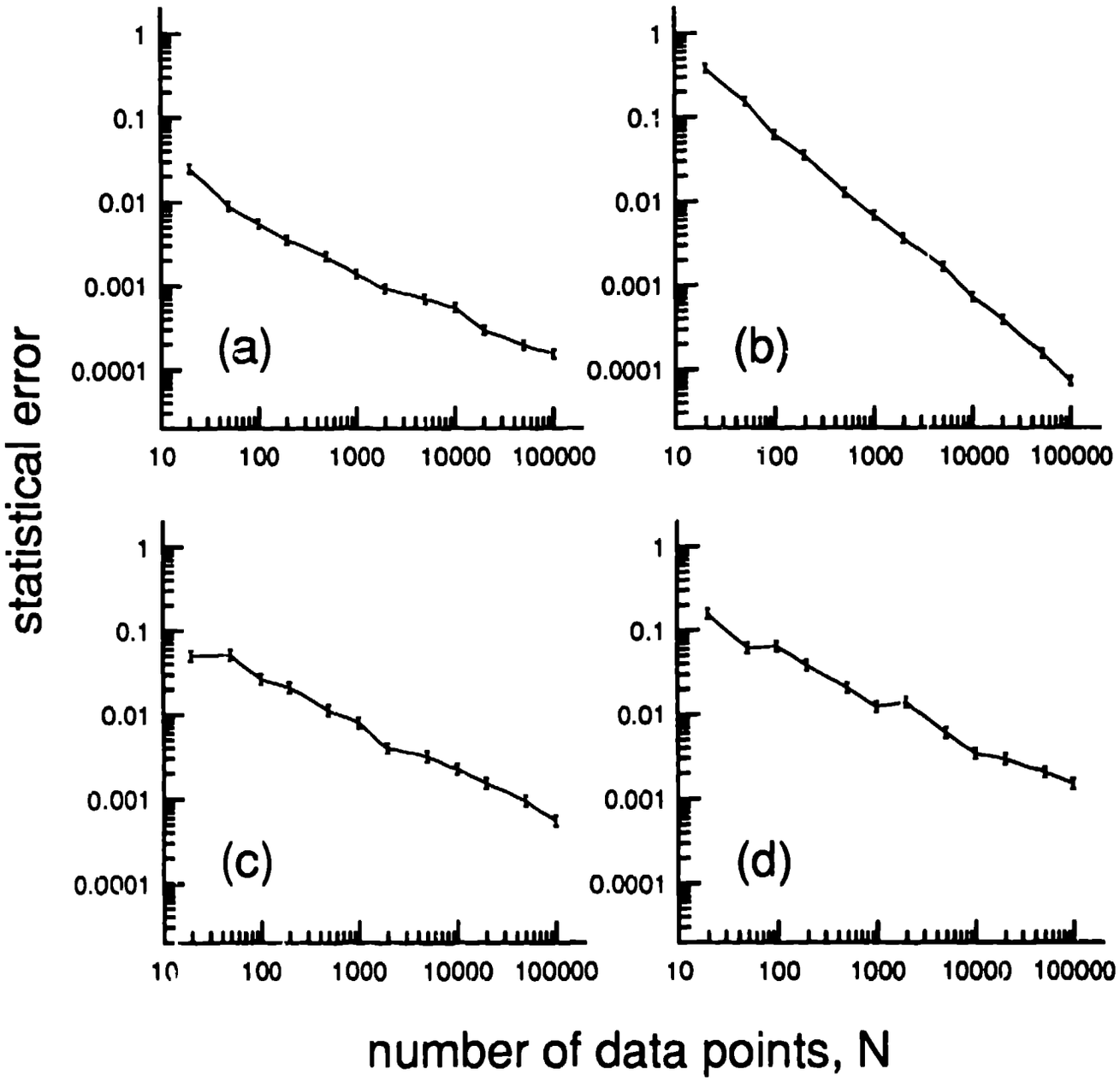

Figure 3: Theiler and Lookman